Leveraging the Potentials of Dedicated Collaborative Interactive Learning: Conceptual Foundations to Overcome Uncertainty by Human-Machine Collaboration

Adrian Calma
University of Kassel
adrian.calma@uni-kassel.de

Sarah Oeste-Reiß
University of Kassel
oeste-reiss@uni-kassel.de

Bernhard Sick
University of Kassel
bsick@uni-kassel.de

Jan Marco Leimeister
University of St.Gallen
& University of Kassel
janmarco.leimeister@unisg.ch

Abstract

When a learning system learns from data that was previously assigned to categories, we say that the learning system learns in a supervised way. By "supervised", we mean that a higher entity, for example a human, has arranged the data into categories. Fully categorizing the data is cost intensive and time consuming. Moreover, the categories (labels) provided by humans might be subject to uncertainty, as humans are prone to error. This is where dedicated collaborative interactive learning (D-CIL) comes together: The learning system can decide from which data it learns, copes with uncertainty regarding the categories, and does not require a fully labeled dataset. Against this background, we create the foundations of two central challenges in this early development stage of D-CIL: task complexity and uncertainty. We present an approach to "crowdsourcing traffic sign labels with self-assessment" that will support leveraging the potentials of D-CIL.

1. Introduction

Advances in automation, artificial intelligence and machine learning are changing our way of working and way of thinking. On the one hand, there is fear that robots will replace the workforce. On the other hand, promising possibilities of human-machine collaboration emerge. This type of collaboration will have the potential to help companies to remain competitive on the market. For example, it has the potential to support companies and human workforce in decision-making processes, help them to develop and offer new intelligent services and products. However, decision making processes are typically influenced by uncertainty and many more factors. Imagine an intelligent system that will support the human workforce in decision making. Thus, an algorithm is needed that copes with uncertainty issues and provides its human collaborators correct information. Therefore, a central basis is constituted by machine learning algorithms that deal with those purposes.

Therefore, in the following we briefly present a motivating case study from us that provides preliminary first results to sensitize for the underlying basic challenges of uncertainty. In the case study students had to label traffic signs that they viewed for a limited amount of time. In the following Section, we succinctly describe the experimental setup and summarize the results of the labeling process.

1.1 Motivating Case Study

1.1.1. Experimental Setup. We preselected 17,400 images of traffic signs from the German Traffic Sign Recognition Benchmark (GTSRB) [1] (the total number of traffic signs is 39,209)], which was proposed in [2]. The preselection of images is motivated by the limited resources, on the one hand, and by our goal to select samples that show greater uncertainty, on the other hand. We aimed at one goal during our preselection: Select the images that are harder to classify with 100% certainty. That is, they either show higher probability to be misclassified or the uncertainty regarding the provided label is high. Consequently, two persons examined all the images in the GTSRB dataset and selected those for which one of them would think that they are hard to classify without any doubt (i.e., the classification is subject to uncertainty).

URI: http://hdl.handle.net/10125/50007
ISBN: 978-0-9981331-1-9

The group of annotators consisted of 7 students, all in possession of a driving license. A labeling session took maximum 20 minutes, with breaks of about 5 to 15 minutes between the sessions. Every student had exactly a total of 7 seconds time to view the image of the traffic sign, select the corresponding class of his choice, to assess the certainty, and to submit his decision. The image of the traffic sign was displayed for one second (this second is contained within the total 7 seconds). After the designated time elapsed, the input fields were blocked, so that the student was restricted from entering any new information. In this special case, the image of the traffic sign was marked correspondingly (tagged as "time's up"). The input fields that were filled in up to this point in time were still saved in the database.

1.1.2 Labeling Outcome. From the total of 17,400 images, 16,567 were labeled correctly by the annotators. From the remaining 833 images, 663 were labeled wrongly, whereas for 170 samples the time elapsed. The total number of images for which the time elapsed sums up to 206: 170 misclassified and 36 correctly labeled. Table 1 and Figure 1 summarize the results of the labeling session. At this point, we would like to emphasize that every image of a traffic sign has been labeled by only one student. We notice that most of the labelers reach an accuracy of about 96%, which is comparable with the human performance of 98.84% on the final and complete GTSRB dataset, as presented in [3].

Labeler	Label	Count	
		#	%
1	correct	9	64%
	wrong	5	36%
2	correct	2975	96%
	wrong	112	4%
3	correct	3006	96%
	wrong	116	4%
4	correct	1710	86%
	wrong	271	14%
5	correct	440	87%
	wrong	67	13%
6	correct	2016	98%
	wrong	41	2%
7	correct	2053	96%
	wrong	81	4%
8	correct	4358	97%
	wrong	140	3%

Table 1: Labeling results of the motivating case study.

Figure 1: Labeling results of the motivating case study. The proportion of the labeled images in the sunburst chart is represented by the size of the inner ring. The number corresponds to the ID of the labeler. The proportion of the misclassified and correctly classified images is represented by the size of the outer slices.

These results support our supposition that, in the future, systems will have to learn from uncertain sources. Figure 2 depicts the certainty distribution over all seven students, where 19.10% of the labels were subject to uncertainty.

This motivates us to set the foundation for handling uncertainty and for designing human-machine collaboration in a dedicated context.

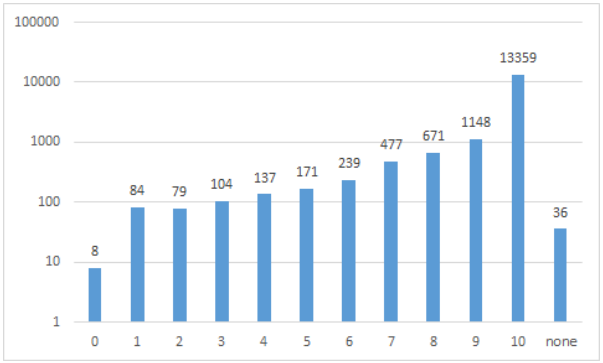

Figure 2: Certainty distribution over all seven students. The lower the value, the higher the uncertainty. The certainty value marked as **"none" refers to the case when the seven** seconds elapsed. The values on the bars depict the number of samples labeled with the corresponding certainty value.

1.2. Problem Statement and Research Questions

Against the before described background, the so called dedicated collaborative interactive learning (D-CIL) [4] seems to be a promising solution. D-CIL is a specific new machine learning paradigm that has the potential to cope with these demands. In a D-CIL context, realistic assumptions are made about the learning task: an annotator (e.g. human domain expert like a crowd worker), generally referred to as an *oracle*, may be wrong or uncertain; there are multiple annotators, with different degrees of expertise which can collaborate to solve the labeling task; and, the learning system provides feedback to the annotators. Besides that, it is important to recognize, the term of 'collaboration'. It refers to the work of two or more actors towards a common goal and has the potential to improve the quality of work products an individual cannot achieve [5, 6]. However, in the context of D-CIL, a group of human experts is needed that helps the system to learn and, in the long run, in order to provide humans with services for decision support. Therefore, a critical success factor might be inherent in the access to human experts as a valuable resource for human-machine collaboration and more precisely for D-CIL. From that point of view, crowdsourcing literature provides additional insights. It deals with outsourcing a task to a group of human experts. In that context, there is an open call (e.g. from a company) in the form of a task that is outsourced to an undefined group of people [7, 8].

Therefore, we base our investigation on these research streams and we answer the following **research questions**:

1) *"What are the conceptual foundations of D-CIL in terms of handling uncertainty"* and
2) "*How should human-machine collaboration in D-CIL context be designed to activate learning mechanisms among a learning system?"*

To answer these questions, we focus on leveraging the potentials of D-CIL by:

- Laying the **foundations** for dealing with **uncertainty** and **task complexity** (see Section 2) and
- Develop a **crowdsourcing solution** as means for developing and establishing human-machine collaboration in terms of D-CIL, to gain insights for concrete **learning mechanisms/ algorithms** (see Section 4)
  - with **multiple** uncertain **oracles** (humans respectively crowd workers)
  - that **self-assess** their uncertainty
  - by participating in a labeling task from an open call from a **crowdsourcing campaign.**

To address the before described research aims, we follow a design science research (DSR) [9] approach. Against that background, our solution makes contributions towards a design theory, since it explains the purpose and scope of D-CIL in terms of reporting conceptual foundations for a learning system. We provide insights for a generalizable crowdsourced solution that helps a learning system to learn and deal with uncertainty.

2. Methodology

The aim of our study is to create the conceptual foundations of a new machine learning paradigm called D-CIL that overcomes the lack of uncertainty. To address this research gap, a socio-technical perspective is needed since human-machine collaboration constitutes a critical success factor. Therefore, the research in our context is more than just developing a learning mechanism/ algorithm. To leverage the potentials of D-CIL, a socio-technical system is needed that incorporates machine learning mechanisms (learning system) and, respects the way of collaboration between humans and machines. For that reason, DSR provides a useful research approach, since it involves the construction of a wide range of socio-technical artifacts like decision support systems [10]. In line with Gregor 2013 [11] we aim to make contributions toward a design theory. In order to achieve our goal, we follow Hevner's three cycle view of DSR [12] (see Figure 3).

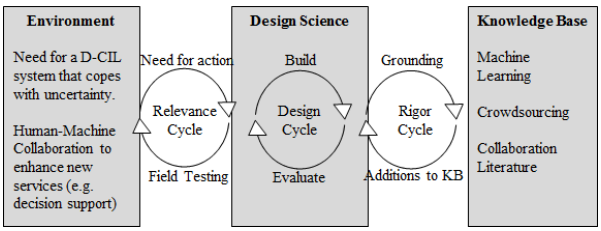

Figure 3: Design Science Approach.

Firstly, we started a relevance cycle by presenting a motivating labeling case study with error prone and uncertain human annotators (see Section 1). Secondly, we started a rigor cycle by drawing on justificatory knowledge of D-CIL, crowdsourcing and collaboration literature and introduce related work (see Section 3). Thirdly, we start a design cycle and present the D-CIL approach in terms of uncertainty, whereas we present its potentials (see Section 4.1) and specify the 'purpose and scope' of our solution inherent in the conceptual foundations of D-CIL (see Section 4.2, 4.3) as well as

'principles of form and function' inherent in the design of the crowdsourced solution for human-machine collaboration (see Section 4.4). Finally, Section 5 concludes the article by presenting limitations and future research work.

3. Related Work

Probably, the most related field to D-CIL is active learning (AL), especially pool-based AL (PAL). In an PAL context, there is a learning entity that has access to a large pool of unlabeled data and a small set of labeled data. Then, in every learning cycle, one sample or a set of samples is chosen from the unlabeled pool and presented to the *oracle*, that provides the correct class information. In a PAL context, the following assumption is made: there is **only one** oracle, who is **omniscient**. A lot of research has been conducted, mostly focused on the selection strategy [13, 14, 15], i.e. answering the following question: *Which is the next most informative sample to be selected for labeling?* It has been showed that, based on the previous restrictive assumption (one omniscient oracle), PAL produces the desired result: performs comparable to *supervised learning* (all data in the pool is labeled) with less labeled samples.

Recently, the fact that labels are subject to uncertainty has drawn the attention of the research community. Still, the research in PAL with *error prone oracles* is in its infancy. Therefore, we point out some research efforts that focused on AL with one error prone oracle:

For example, the oracle can be asked to provide a confidence level for its answer (e.g. in binary classification problems), whereas the selection strategy handles the trade-off between maximizing the information (in this case the entropy) of a sample and minimizing the probability that the oracle will be unconfident [16]. A selection strategy for AL on binary data has been presented in [17]. It is based on two assumptions: (1) the higher the confidence of the oracle, the more likely that the answer is correct and (2) the higher the confidence of the learning system, the more likely the oracle is too. Thus, a trade-off between exploring the unlabeled data and exploiting the labeled data is proposed.

Two further approaches (for multiclass problems) were proposed in [18]: The first one, *Disagreement 1* measure the "influence" of a sample by determining the disagreement between a model learned from labeled data and one learned from data labeled by the first one. Concretely, the goal is to find the sample that influences the model the most. *Disagreement 2*, on the other hand, aims at identifying samples that are incorrectly classified by the learning system [18].

Yet another idea is to "forget" the labels for the samples that are responsible for increasing the error level in the learned model [26].

Up to this point, we presented related research efforts that consider one uncertain oracle. But, research has been conducted with more than one uncertain oracle, too: A strategy to handle the trade-off between re-labeling and single labeling has been proposed in [19]. A different approach is adopted by the STAL framework [20]: the learning system determines the oracle that is most reliable for the sample to be queried. Moreover, the most unreliable oracle learns from the most reliable one.

The strategy ALJ [21] goes one step further and estimates not only the labels and the oracle`s expertise but also the difficulty level of a sample in the context of crowdsourcing. One further approach that focuses on crowds assumes that there exists an omniscient oracle [22]. First, the data is labeled during crowdsourcing and then the labels are inferred from a specific algorithm (e.g. [23], [24], [25]). Subsequently, labels for samples most likely to be labeled incorrectly are queried from an omniscient oracle.

In relation to existing related work in this field, it is important to delineate D-CIL to the existing paradigms and refer to open research opportunities. In the previously presented, similar approaches, there is no bidirectional interaction between the oracles and the learning system, the oracles label only samples, and there is no collaboration between the oracles. But, in a D-CIL context,

- the learning system provides feedback to the oracles,
- the oracles may evaluate rules generated by the learning system, and
- the oracles collaborate with each other (except for [20]).

4. D-CIL Approach in Terms of Uncertainty

4.1. Potentials of D-CIL

As we pointed out in Section 2, D-CIL is more than just a learning mechanism or algorithm. It opens a socio-technical system perspective. Therefore, we refer in the following to mid-term potentials of D-CIL to delineate its scope and transfer it to economical contexts. Consider the following practical problem: Cars get broken. Thus, the owners drive to a car service to let the car be repaired. But, the same malfunction or symptoms are encountered by other car

owners too. Therefore, the car manufacturer might be interested in creating a car diagnostic system to save time for troubleshooting and money. For this reason, it has to store the provided solutions in a database for being able to address it later.

Figure 4 depicts schematically this example: There is large set of car problems and their descriptions and, generally, a significantly smaller set of car services. Every time car service is confronted with a car problem it will deal with it; ideally it will fix it. Still, we can assume that some solution will be provided. As the same car issues may appear simultaneously the car services might provide different solutions for the same issue. These information, the pairs consisting of problem description and solution, are added to a *knowledge database* which can be queried by the diagnose system.

As a new car model is released, the diagnose system should be able to learn from scratch. A possible design solution is presented in Figure 5: The diagnose system must be able to determine the order in which the car problems are dealt with. That is, it must have access to the pool containing the descriptions of the car problems. Furthermore, its decision is based on an appropriate selection strategy, that will select the next *most informative* car problem. By *most informative* we mean the car problem, that when solved, will bring the maximal gain. Then, the car service is requested to deal with selected issue. Of course, the car itself, is only presented in a car service. But, the description of the problem can be send to other car services too. The different solutions are then aggregated and the diagnose system updated. Thus, we will able to *learn* by being curious (asking questions, e.g. selecting the data from which we learn) and reasoning (aggregating the provided solutions).

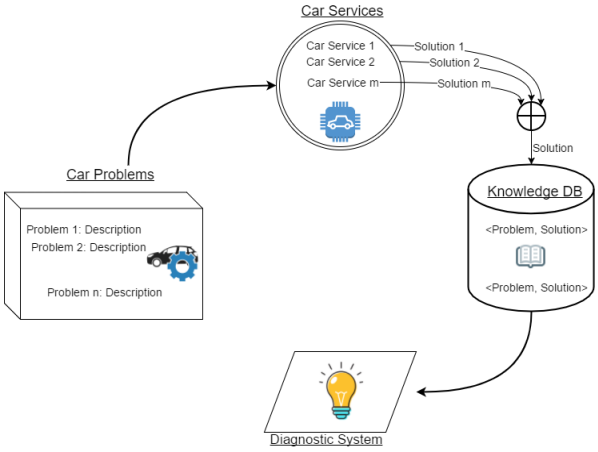

Figure 4: An Example of a possible Learning Problem.

The diagnose system, the selection strategy, the aggregation strategy, and the knowledge database form a *learning system* (depicted in Figure 5), which exhibits the following abstract properties:

Curiosity: It selects the data from which it learns (by means of the selection strategy) and

Reasoning: it can deal with multiple, sometimes contradictory and uncertain, information (by means of an aggregation strategy).

A step in this direction was taken by conducting a case study with image data on an apparently simple classification task, to inspect how humans self-assess their uncertainty, as presented in Section 1.1.

4.2 Guiding Idea of D-CIL

In the following, we describe the guiding idea of D-CIL in more detail and refer to its core characteristics. D-CIL can be described as a socio-technical machine learning approach that bases on the collaboration of humans and machines as well. To refer to the terms of D-CIL [4], they can be described as follows:

- **Dedicated**: The learning task is clearly defined and the number oracles is relatively small.
- **Collaborative**: The oracles (e.g., human domain experts) collaborate to provide the information.
- **Interactive**: The information flow is bidirectional: from oracles to the learner and vice versa in form of feedback.

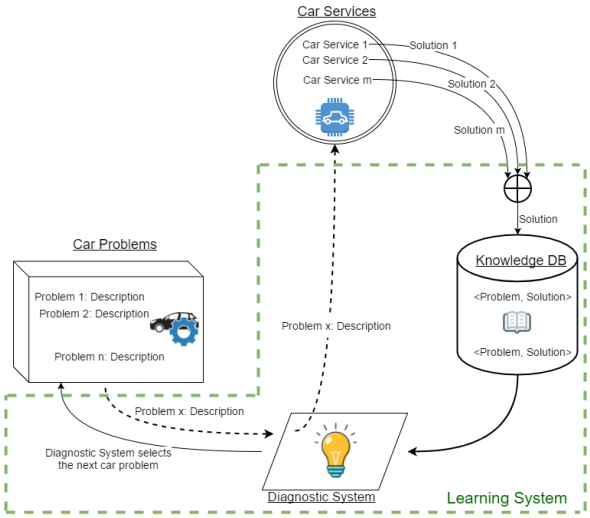

Figure 5: Motivating Example.

More generally, it can be described as a learning cycle (sketched in Figure 6): We have access to a *large pool U of unlabeled data* and we can ask different entities – such as humans, simulation systems, or test stands – that can communicate with each other for

additional information, i.e., for labeling. We will address these entities under the general term of *oracles*. But, as these systems are not omniscient, we must assume that the provided information – i.e. the labels – may be erroneous. However, the *learner* can aggregate the information and add it to a relatively *small pool L of labeled data*. A *knowledge model* (e.g. a decision entity) is constructed based on *L*. The data in *U* is evaluated by means of a *selection strategy* and the *most informative* data is selected for labeling. Over time, as the *learner* has a solid knowledge model, it will provide *feedback* to the oracles, too.

4.3 Conceptual Foundations of D-CIL

In the following, we refer to the conceptual foundations of D-CIL in more detail.

What do we mean by "uncertain"? In [27], *uncertainty* is used as a generic term for addressing aspects such as "unlikely", "unreliable", "imprecise", or "vague". When humans are asked to provide information about an actual situation, the confidence regarding the given answer depends on diverse factors, such as the difficulty to assess that information, previous experience, or knowledge. Certainly, there are times when we cannot state our answer with absolute confidence. Thus, we tend to add additional information about the quality of our answer, i.e., to quantify and qualify our confidence [27].

What are possible reasons for uncertainty? The performance of humans depends on different factors such as experience, expertise, concentration, or fatigue level.

The difficulty of a labeling task is given by the number of the steps the annotator should perform in order to determine the right class, the knowledge required for understanding the problem, the experience with similar labeling tasks, the designated time, and the risk involved by a misclassification. For example, if we are presented with a picture and asked, "is there a cat in the picture?", we might have a less complex task to fulfil. Still, our answer depends on how we interpret the notion of *is there*: if we only see the tail, will we answer positive? Furthermore, it assumes some knowledge: we know what a cat is. How will we answer, if we are shown a picture with a lion or the picture of a liger? An example of a complex classify-cation task is deciding if a patient must undergo surgery. This, usually, involves performing thorough analysis by multiple qualified personnel, thus the decision is based on heterogenous information sources. Moreover, the risks/costs involved by deciding against a surgery when the patient needed one are higher than the other way. In addition, the decision has to be taken under time pressure (e.g. emergency operation). A further source of uncertainty is the lack of ground truth, which is missing because it is impossible (e.g. will a car break down in the next two years?) or too expensive to assess at time of labeling (e.g. which of the five prototypes will sell best?).

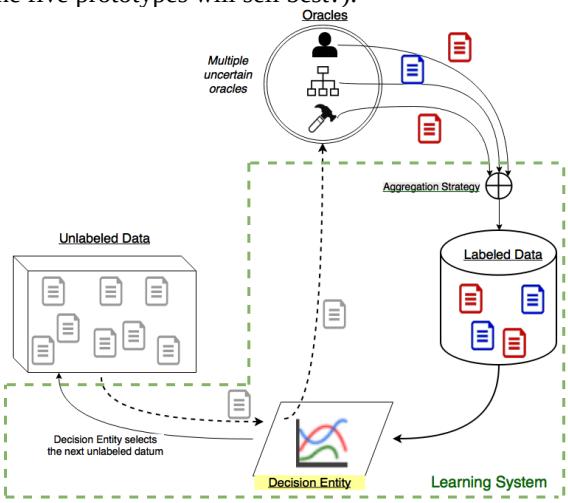

Figure 6: Learning Concept in a Dedicated Collaborative Interactive Learning Setting.

Against that background we derive the following general assumptions that guide our idea:

- **Assumption 1:** At the beginning of the learning, the learner has access to a large set of unlabeled data. This data set is either free or can be purchased at low costs;
- **Assumption 2:** For any data point in the data set we can buy additional information, i.e. labels.
- **Assumption 3:** The costs for acquiring labels are uniformly distributed, i.e. the costs are the same no matter which oracle we address or which data point we select for labeling.
- **Assumption 4:** The oracles are prone to error; thus, the labels are subject to uncertainty.

In the following, we therefore focus on how we extract knowledge from uncertain oracles.

4.4 Crowdsourced Human-Machine Collaboration to Overcome Uncertainty

Task and Context Specification: To investigate to which degree we can extract knowledge from uncertain oracles we need a possibility to evaluate the performance of the learning system. We have decided to address a classification task, as it is straightforward to evaluate the performance of a classifier (e.g. accuracy, confusion matrix). Thus, we selected a data set for which we know the true labels. We decided in favor of an image data set, the German Traffic Sign

Recognition Benchmark [1] as any person with a driving license may be considered a domain expert and the ground truth is available.

As shown in [3] and confirmed by our motivating case study (Section 1.1), the humans could label 98.84% of the traffic signs correctly. But we are interested in data which exhibits a higher degree of uncertainty, in order to simulate harder learning tasks. We decided to manipulate the data by applying blur filters, changing the brightness, or blackening an area of the image. We justify our decision by different light conditions, snow or tree branches, which may influence the visibility of the traffic signs. Furthermore, to add even more uncertainty we can limit the display time.

This approach provides the potential to *develop a learning system that can learn from uncertain oracles*. The essential advantage is that we can determine whether the system is learning correctly or not. This step is required for being able to deal with more complex tasks which are truly subject to uncertainty. Summarized, we have set the following goals for the labeling process:

- Integrate error prone humans (i.e. uncertain oracles) into the learning process;
- Reduce cognitive load for human experts by carefully designing the labeling process, and
- Gaining insights into human motivation and collaboration with the learning system.

Procedures of the Human-Machine Collaboration: Guided by task and context specification, human experts are needed to solve those tasks. On the one hand, the procedures of solving the task need to be designed in a reusable and systematic manner. On the other hand, access to human experts is needed.

Therefore, we developed a learning system that includes a designed reusable process that supports human-machine collaboration to solve a labeling task. We use a crowdsourcing campaign to get access to human experts in the form of crowd workers. Therefore, a crowdsourcing campaign will be conducted on Amazon's Mechanical Turk. We will provide an open call for a labeling task. The oracle, in this case the crowd worker, will be forwarded to our learning system, where it starts the human-machine collaboration.

The procedures of the human-machine collaboration can be described as a reusable process. The intention of the process design is to achieve correct solutions from humans. To achieve correct solution, the human-machine collaboration should minimize cognitive load. Overall, the humans will **see the image** of a traffic sign for **a limited time**. They will complete a 4-step labeling procedure.

- **Step 1:** The oracle will have to choose between three categories: *round*, *square*, or *indecisive*.
- **Step 2:** The further labeling process depends on what the oracle has labeled in the previous step. Suppose he has selected:
  - *round*, then the oracle has to choose from further four categories that best describe the observed image: *red*, *blue*, *black/white*, or *indecisive*.
  - *angular*, then the oracle may choose between *triangle*, *other*, or *indecisive.*
- **Step 3:** Depending on the choices made in the previous two steps, the oracle will see sample images of the traffic signs and must make the final selection. For example, if the oracle has previously chosen *round→red*, then, in the final step, he can choose between speed limit & prohibition signs (e.g., no passing sign). Similarly, depending on the previous selections (e.g., *round→blue*, *round→black & white*, *angular→triangle*, etc.) the corresponding images of the sample traffic signs are presented for selection. If at any point the oracle selected *indecisive*, all sample traffic sign images will be presented to it.
- **Step 4:** This is probably the most important step, as the oracle must self-evaluate its own certainty. He should fill in a value between 0 and 9 (i.e., the evaluation scale has a precision of 10), which represents the self-assessment.

We can assume that in case of a real-world problem, there might be time constraints that will require the user to respond immediately. Thus, we track the time an oracle needs for the labeling process. That is, we track the time elapsed between the moment the image was shown and any interaction with the labeling system. By doing so, we can simulate situations in which the learning system receives only partial input. It helps us develop mechanisms that can manage different degrees of missing information and different levels of response times. Moreover, we may investigate if there is a correlation between the time necessary to completely label the data points and the (un)certainty.

5. Limitations, Future Research, Contribution, and Conclusion

The presented approach is conducted on image data, which may be seen as a severe limitation, but considering the very early stage of research in the area of D-CIL, it is necessary to start with data that we can easily understand. For the same reasons, we do not address collaboration between oracles and feedback from the learner to the oracles. Another limitation is

the fact that we perform a simulation, but this is vital in this early phase of development. It allows us to develop, investigate, and evaluate techniques for selection and aggregation strategies, for collaboration methods, and for providing qualitative feedback, with the ultimate goal of bringing D-CIL to practice. Possible feedback may include a quantitative report about the individual performance compared to the other oracles, about its individual failure rate, a peer assessment report based on numeric grades by other human oracles, or a peer assessment report based on review criteria and a textual review. Moreover, the learning system reveal previously labeled samples that are similar to the current one, but which have been labeled differently.

The next step is to develop suitable techniques for selecting the next *most informative* data point (e.g. traffic sign image) that should be presented to the oracles for labeling. Addressing the challenge of dealing with *uncertain*, in some cases even contradictory, information provided by the oracles enjoys the same importance as the selection strategy. Additionally, uncertainty may be induced by lack of information: If the time had expired, before the oracle finished to fill in all the input fields, then we end up with a partial answer. Thus, we must deal with this kind of uncertainty too.

Furthermore, we may want to consider prior knowledge. For example, we have the large set of traffic sign images but we do not know which traffic signs they represent (i.e. we do not know to which category they belong to). But we have access to representatives from each category, i.e., we know how the traffic signs should look like. Thus, we can harness the potential of prior knowledge: we can apply machine learning techniques that will extract features ([27, 28]), such as the predominant colors or if an image contains numbers or not. This helps us cluster (an active learning paradigm for clustering is presented in [30]) the unlabeled data which might reduce the human labelling effort.

Obviously, the goal is to bring D-CIL into practice. Therefore, we aim at conducting a D-CIL experiment with data for which the ground truth is missing at the labeling time.

In this article, we created the foundations for addressing the challenge of uncertainty by presenting an approach to *crowdsourcing traffic sign labels with self-assessment*, which leverages the potentials of D-CIL.

Thus, we could make a first step toward making D-CIL common practice in situations where the ground truth is missing.

6. Acknowledgements

This paper presents research that was conducted in the context of the project “CIL” (funding program for further profiling of the University of Kassel 2017-2022: “Zukunft 2017-Standard”).

7. References

[1] Institut für Neuroinformatik, German Traffic Sign Recognition Benchmark. http://benchmark.ini.rub.de/?Section=gtsrb (2017) [Online; last accessed 2-September-2017]

[2] Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C., “The German Traffic Sign Recognition Benchmark: A multi-class classification competition.” In: IEEE International Joint Conference on Neural Networks, San Jose, CA (2011), pp. 1453–1460

[3] Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C., “Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition.” Neural Networks, Vol. 32 (2012) , pp. 323–332

[4] Calma, A., Leimeister, J.M., Lukowicz, P., Oeste-Reiß, S., Reitmaier, T., Schmidt, A., Sick, B., Stumme, G., Zweig, K.A., “From active learning to dedicated collaborative interactive learning.” In: 29th International Conference on Architecture of Computing Systems, Nuremberg, Germany (2016), pp. 1–8

[5] Briggs, R. O.; Kolfschoten, G.; de Vreede, G.-J.; Douglas, D.: “Defining Key Concepts for Collaboration Engineering.” 12th Americas Conference on Information Systems (AMCIS), Acapulco, Mexico 2006.

[6] Briggs, R. O.; Kolfschoten, G. L.; de Vreede, G.-J.; Albrecht, C.; Lukosch, S.; Dean, D. L. (2014): A Six-Layer Model of Collaboration. In: Collaboration Systems. Hrsg.: Jay F. Nunamaker Jr.; Nicholas C. Romano Jr.; Briggs, R. O., Advances in Management Information Systems New York 2014, pp. 221-228.

[7] Durward, D.; Blohm, I.; Leimeister, J. M. (2016): Crowd Work. In: Business & Information Systems Engineering, (2016), pp. 1-6.

[8] Zogaj, S.; Leicht, N.; Blohm, I.; Bretschneider, U. “Towards Successful Crowdsourcing Projects: Evaluating the Implementation of Governance Mechanisms.”In: 36th International Conference on Information Systems (ICIS 2015), 2015, Fort Worth, TX.

[9] A. Dresch, D. P. Lacerda, J. A. V. Antunes Jr, Design Science Research, Springer International Publishing, 2015

[10] Gregor, S.; Jones, D., “The Anatomy of Design Theory.” In: Journal of the Association for Information

Systems (JAIS), Vol. 8 (2007), No. 5, pp. 312-335.

[11] Gregor, S.; Hevner, A. R., "Positioning and Presenting Design Science Research for Maximum Impact". In: MIS Quarterly (MISQ), Vol. 37 (2013), No. 2, pp. 337-355.

[12] Hevner, A. R., "A Three Cycle View of Design Science Research." In: Scandinavian Journal of Information Systems, Vol. 19 (2007), No. 2, pp. 87-92.

[13] Constantinopoulos, C., Likas, A. C.: "An incremental training method for the probabilistic RBF network." In: IEEE Transactions on Neural Networks, Vol. 17 (2006), pp. 966–974

[14] Donmez, P., Carbonell, J.G., Bennett, P.N.: Dual strategy active learning. In: 18th European Conference on Machine Learning, Warsaw, Poland, Springer (2007), pp. 116-127

[15] Guo, Y., Greiner, R.: Optimistic active learning using mutual information. In: Proc. of Int. Joint Conf. on Arti. Intell., Hyderabad, India (2007), pp.823-829

[16] Fang, M. and Zhu, X.. 2013. Active learning with uncertain labeling knowledge. Pattern Recognit. Lett. 43 (2013), pp. 98–108.

[17] Jun Du and Charles X. Ling. 2010. Active learning with human-like noisy oracle. Proc. - IEEE Int. Conf. Data Mining,ICDM (2010), pp. 797–802.

[18] Mohamed-Rafik Bouguelia, Slawomir Nowaczyk, K C Santosh, and Antanas Verikas. 2017. Agreeing to disagree: active learning with noisy labels without crowdsourcing. Int. J. Mach. Learn. Cybern. (2017), pp. 1–13.

[19] Christopher H Lin and Daniel S Weld. 2016. Re-active Learning: Active Learning with Relabeling. AAAI (2016), pp. 1845–1852.

[20] Meng Fang, Xingquan Zhu, Bin Li,Wei Ding, and Xindong Wu. Self-Taught Active Learning from Crowds. (2012).

[21] Liyue Zhao. 2014. An Active Learning Approach for Jointly EstimatingWorker Performance and Annotation Reliability with Crowdsourced Data. ArXiv (2014)

[22] Zhenyu Shu, Victor S Sheng, and Jingjing Li. 2017. Learning from crowds with active learning and self-healing. Neural Computing and Applications (2017), pp. 1–12.

[23] Victor Sheng, Foster Provost, and Panagiotis Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008), pp. 614–622.

[24] Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning from Crowds. Journal. Of Machine Learning Research Vol. 11 (2010), pp.1297–1322.

[25] Jing Zhang, Victor S. Sheng, Jian Wu, and Xindong Wu. Multi-Class Ground Truth Inference in Crowdsourcing with Clustering. IEEE Transactions on Knowledge and Data Engineering, Vol. 4 (2016), pp. 1080–1085.

[26] Xiao Yu Zhang, Shupeng Wang, and Xiaochun Yun. Bidirectional active learning: A two-way exploration into unlabeled and labeled data set. IEEE Transactions on Neural Networks and Learning Systems, Vol. 12 (2015), pp. 3034–3044.

[27] A. Motro and P. Smets, Eds., Uncertainty Management in Information Systems – From Needs to Solutions. Springer US, 1997.

[28] R. F. Murphy, "An active role for machine learning in drug development," Nature Chemical Biology, vol. 7, pp. 327–330, 2011.

[29] J. D. Kangas, A. W. Naik, and R. F. Murphy, "Efficient discovery of responses of proteins to compounds using active learning," BMC Bioinformatics, vol. 15, no. 143, pp. 1–11, 2014

[30] R. Marcacini, G. Correa, and S. Rezende, "An active learning approach to frequent itemset-based text clustering," in Int. Conf. on Pattern Recognition, Tsukuba, Japan, 2012, pp. 3529–3532.